SEON

seon.io

Guide to Fraud Analytics in 2022

Table of Contents

One of the cornerstones of any good fraud prevention strategy is the use of data analytics.

By looking at past data with analytical methods, we can single out characteristics that are more likely to be fraudulent – and can devise rules, measures and procedures to guard our business against it.

The main advantage of using analytics over just looking at single incidents is that any insights gained during the analytical process can be used to scour through the system and uncover potentially harmful cases that might have been missed initially.

This guide will walk you through the principles of fraud analytics, allowing you to uncover fraud patterns that would otherwise remain hidden.

By investing in the analytical approach, you can save money in the long term by preventing fraud that otherwise would've turned into straight losses before they materialize.

What Is Fraud Analytics?

Fraud analytics is the use of data analytics in the context of fraud prevention.

This ranges from simply looking at available data to find anomalies, as signals outside of expected/normal ranges are considered suspicious or risky, to big data techniques and machine learning to automatically detect and prevent fraud at scale.

This last point is crucial: Any online service or business will have to rely on analytics to fight fraud, as the number of transactions and events that go through a system is simply too big for humans to analyze manually.

Fraud analytics is the methodological power tool to **quickly uncover areas of interest**, and the use of statistics allows for setting up rules to **prevent undesirable events** as well as aiding system and service design to **minimize risk**.

The main approach is combining both business intelligence data and other internal data on customer actions and transactions and identifying patterns in order to build up models.

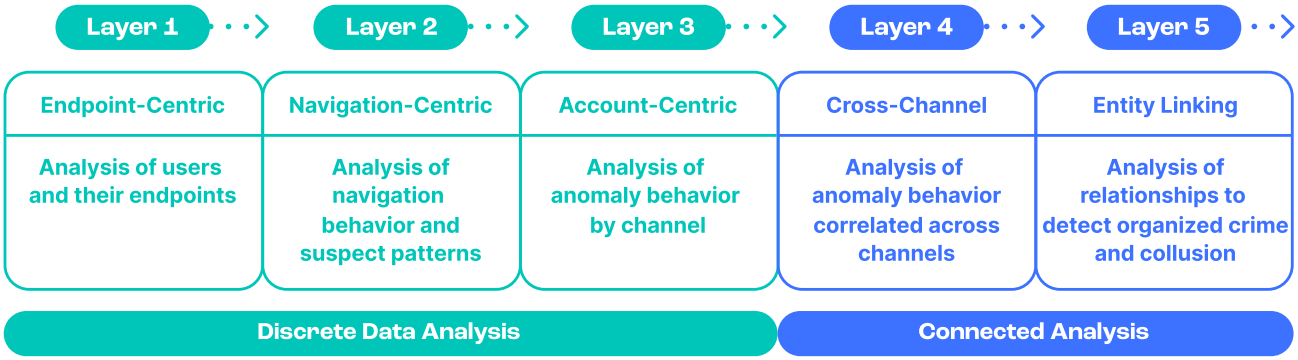
As a business evolves, it will have more and more data available, often siloed in different systems, and fraud analytics needs to bring these different data sets into one pool in order to make meaningful connections.

How to Use Data Analytics for Fraud Detection

Experts recommend using a layered approach to use data analytics for fraud detection, based on what data is available at different stages and levels of aggregation.

The idea here is that using data analytics, we can detect anomalies at different levels and with different granularity, making it real hard for even the best attacks to get through undetected.

Gartner's Layered Fraud Prevention Approach



These five layers, as defined by Gartner, are:

- endpoint centric: analyzing users at their endpoints (devices, registration information, etc.)
- navigation/behavior centric: looking at service usage data to detect anomalies/outliers
- account centric: looking at the channel level for suspicious behavior (e.g. by user cohorts, markets)
- cross channel: looking at how anomalous behavior correlates in different channels (eg. comparing organic users with affiliate program users)

- entity linking: in-depth analysis of connections between different entities in the system to uncover complex attacks

This is helpful in thinking about how to layer different levels of data aggregation together to run analysis on. Think of it as comparing apples to apples in order to detect an orange.

By aggregating data on these levels, we can easily set up triggers for activity that lies outside what's statistically expected to happen, in a way that makes sense in the given context.

For example, if we wanted to detect fraudulent or bad signups,

1. We would first look at how our good users looked at their signup stage (in a given market or otherwise well-defined segment).
2. Then we would look at how bad users (those resulting in chargebacked transactions for example) looked like.
3. We'd compare the two cohorts to find differences.
4. After analysis, we could identify certain differences that can be translated to all five layers of data aggregation, where the bad users are statistical outliers compared to the good ones.

Using this information, we can now come up with rules and triggers to alert us of possible risks not just when we see new users who fit the model of the "bad" group – but also to raise suspicion when we see something that's not clearly in the "good" group.

The way to do this ranges from crude discovery via custom-built BI dashboards to applying data science models on large datasets.

The important thing to keep in mind is to **make sure you're zeroing in on data that's comparable**, and to have a good idea of **how the different stages map on to one another.**

What might not be risky on one stage might be on another and, vice versa, what looks rosy on one level might be problematic on another. As long as your approach

is systematic, you should be well protected without being too strict.

Data Analytics Techniques for Fraud Detection

There are different kinds of data analytics techniques that you can use to analyze your data using statistics. There is no one-size-fits-all approach, and these should be used in conjunction with one another.

Similar to the philosophy of the layered approach, different methods will yield different results – covering as many angles as possible to detect fraud.

Unsupervised Learning (Descriptive Analytics)

These describe statistical methods to automatically discover outliers in your data set to find fraud or risky behavior.

1 Statistical outlier detection method:

- **Standard score/z-score:** calculating the given value's distance from the mean of the entire value set can in itself become a risk score.
- **Break point:** looking at sudden changes in behavior for the given entity
- **Peer group analysis:** sudden changes compared to peers
- **Recency, frequency, monetary value:** compare if a given entity differs from the expected RFM
- **Association rule learning:** discover hidden associations between different variables using if, then rules

2 Clustering algorithms

There are many kinds of clustering algorithms available, and they all aim to group a set of objects based on their similarity to one another in some sense and their differences to others.

When it comes to fraud detection, you have a plethora of data points available for this kind of clustering, and this allows you to discover fraud patterns that a human analyst might miss.

Social Network Analysis and Fraud Graphs

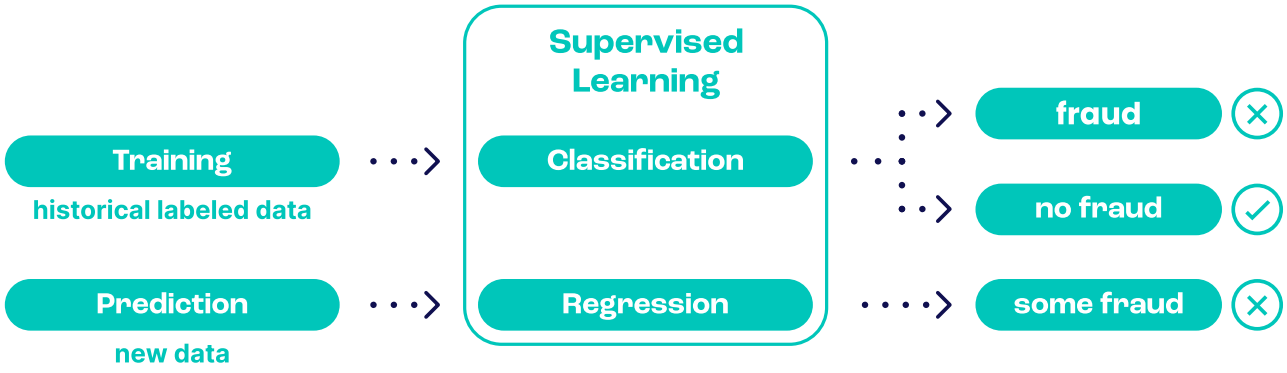
Birds of feather flock together as the saying goes, and it's doubly true for fraud.

Social network analysis means looking for user connections either within your system based on data points (devices, IP addresses, phone numbers, emails and many more), as well as using OSINT to map out connections outside of your system between your users.

Organized criminal gangs who have poor operational security will likely have traces online of their real-world connections.

Predictive fraud analytics

This is the use of historical data and statistical modeling to predict future outcomes.



How Predictive Analytics Supports Anti-Fraud

Applying analytic tools on your historical data allows you to create predictive models that can accurately flag suspicious behavior by applying risk scores, and if the score is above what's accepted, automatically decline a signup or a transaction.

There is a simple, five-step process to follow in implementing this.

1 Labeling your historical database

For this you can also use a sample but, if you can, you should use all your available data. It's simply classifying historical actions and transactions as good or bad (fraud or not), to serve as input for your analytics or learning models.

With SEON, you can use the Label API to do this easily, using transactions that turned out to be fraudulent to train the machine learning model. It will then automatically suggest rules based on the data provided.

2 Fraud pattern & variable analysis

Not all variables are equal, and this step serves as a sanity check. Disregard information that is distributed so equally that it doesn't serve as a meaningful indicator, as it will only add noise to your model.

3 Data modeling

Using the methodology provided above you can now find outliers or risky actions that you haven't previously considered. With these new insights, you should be able to pinpoint what are the strong indicators of fraud, allowing you to set up rules against it.

4 Model implementation

Within SEON this is done either by setting up new rules based on your findings or using the filtering function to discover transactions in your system that fit your criteria in order to uncover fraud.

5 Monitoring & feedback

Over time, your implemented models might become outdated, while fraud can change to escape detection.

This is why you need to monitor your results from time to time and make adjustments accordingly. Again, the Label API can come in handy to update the system in accordance with new discoveries, adjusting for false positives and negatives.

Data Analytics Solutions for Fraud Prevention

There are a myriad of tools available to choose from when it comes to data analysis. The question is, what goals you are looking to achieve?

Approach A: Dashboards

The first approach is building dashboards for your departments that display the aggregated data in a way that it's easy to find anomalies. For this, you can use Tableau, Google Data Studio or other reporting software, which your BI team can help you assemble.

Ideally, setting up these dashboards will allow you to have **an aggregated view of activities**, which will make it easy to conduct audits as well as routine checks. Remember, fraud can slip through the cracks of one layer of detection, only to show up as unexpected behavior on a different level of abstraction.

This approach shields you from both the problem of getting lost in the details as well as countering data silo problems that tend to form on a departmental level.

Approach B: Human-Driven

The second approach would be relying on your data science team, who will have their own favored environments and tools, such as R or Python pandas. They can be relied on to build their own models based on historical data, working in conjunction with your anti-fraud team towards common goals.

Here, SEON's data enrichment of IP addresses, emails and phone numbers can uncover hidden connections in the dataset by adding more variables that can be meaningful – such as fraud rings that turn out not to have social media presence, or accounts that are connected via their email naming convention patterns, and so on.

Approach C: ML-Driven

The third way is letting **artificial intelligence** do the analysis for you.

SEON's whitebox machine learning module, if trained properly through the Label API, will recommend rules automatically based on what patterns it deems risky. The reasoning is transparent and human readable, and it comes with a rule tester so you can determine if it's accurate or not. You can read more about how it works **under our documentation**.

SEON

SEON Technologies Ltd.
seon.io

info@seon.io
+44 20 3997 6090